

# Dictionary of Common Terms

Below is a living set of terms used within EDH. Consider this a glossary or lexicon of the language used in EDH discussions.

NOTE: These terms are put in the order they were created. To reorder them, click the "Term" heading below.

Term	Definition
Activity	A representation of an external datapoint that may or may not relate to a person within EDH. <b>Example:</b> a marketing activity might be a seminar at a hospital, or a clinical activity might be an ER visit.
Addressing	See "Address Preferencing"
Address Preferencing	The process by which each unique person at an address is given a priority number. This number is used to determine who should receive a direct mail when multiple persons at that address appear on the same list. The idea is that we can save money by not mailing the same address twice with the same direct mailer.
AID	A unique identifier given to each unique address. See "Householding".
Baldur	The application used to import data into the EDH database. Conceptually, it is made up of the jobs used on the source data such as cleanse, enrich, collapse, etc. The CLI is used to execute Baldur jobs.
Cassandra	Apache Cassandra, open-source distributed NoSQL database. The EDH data processes initially interact with the Cassandra data store.
Cleanse	EDH logic that standardizes and normalizes information received from input sources. An example includes the Anchor address.
CLI	Within EDH, the Baldur Command Line Interface. It is comprised of many commands used to interact with the EDH database. More information on using the CLI are available at <a href="#">Using the EDH CLI</a> .
Client ID	See "Customer ID". This term is commonly used by applications outside the EDH but still internal to Influence Health.
Control Group	See "Holdout".
Customer ID	The unique identifier that is used internally within the EDH
DNS	Stands for Do Not Solicit. This term is used to mean a list of individual persons that should not be contacted via the given channel. Most often this is used for mail campaigns.
EDH	Enterprise Data Hub, the data layer of Influence Health's CRM.
Elasticsearch	Open source search engine. ETL processes sync data from Cassandra to Elasticsearch for availability to Audience Insights for list builds.
Email Suppression	This is another term for a DNS (Do Not Solicit) list, however this is specifically for the email marketing channel.
EMR	Amazon EMR provides a managed Hadoop framework that makes it easy, fast, and cost-effective to process vast amounts of data across dynamically scalable Amazon EC2 instances. You can also run other popular distributed frameworks such as <a href="#">Apache Spark</a> , <a href="#">HBase</a> , <a href="#">Presto</a> , and <a href="#">Flink</a> in Amazon EMR, and interact with data in other AWS data stores such as Amazon S3 and Amazon DynamoDB.
Enrich	EDH jobs used to improve the data received from external sources. This includes actions such as standardization and normalization of the data needed prior to combining the information with data from other sources.
Experian	Experian plc is a consumer credit reporting agency. Experian collects and aggregates information on over one billion people and businesses including 235 million individual US consumers and more than 25 million US businesses.
Facility Location	The name of a specific hospital/clinic building.
Fixture	A small set of data that helps model a specific condition in the data. These fixtures are commonly used to demo or test development work.
HHID	The last name given to a household. Stands for Household ID. See "Householding".

Holdout	A person that has been randomly set aside from a list of ideal targets to be withheld from any marketing campaigns. The person set aside in this way is someone that would normally be included in the list but because they were randomly chosen will be excluded. This term is most often used in direct mail campaigns. A holdout is used when reporting the ROI on a specific campaign. Essentially, the holdout group replicated what might have been the response if the campaign never actually happened. When you compare the group that were marketed to against the holdout group, you can understand how effective your marketing campaign truly was.
Householding	A process by which an ID is given to each unique address (See "AID") as well as assigning a single last name (See "HHID") for all the individuals that live at that address.
Kafka	Apache Kafka, open source stream processing platform, massively scalable pub/sub message queue architected as a distributed transaction log.
Lead	<p>Depending on the context this could two slightly different things. If the context is more technical, it is referring to a specific piece of data about an individual person that comes from the ULM (Unified Lead Management) application. More broadly a lead could refer to an instance where a person has interacted with a marketing channel.</p> <p><b>Example:</b> This could be a form that a person filled out with their information, or it could be a link that someone clicked on in a social media feed.</p>
Location Preference	See "Location Scoring"
Location Scoring	<p>The location or locations that is assigned to each individual person based on the business rules for location scoring. This score is used to determine which location the person prefers or is closest too.</p> <p>EDH logic that determines the best location to market to a person based on their location and previous interactions with a health system. This is used for organizations that have more than one location being sponsored in marketing campaigns.</p>
Market Area	See "Service Area" the only difference is that this is used more commonly by clients to describe their own service area. Also it is possible for the client to call each grouping under the "Service Area" something other than, primary, secondary, and tertiary.
Person	<p>The unique representation of an individual and all of the latest information that IH knows about them.</p> <p>Types include: New Mover ( n), Prospect (p), Family Member (f), Qualified Prospect (q), Patient (c)</p>
Person Collapse	EDH logic that compares person records to identify matches and relate information that is received from disparate sources to combine the information for a more holistic view of the person.
Propensity	The likelihood that an individual person will need a specific procedure or service from a hospital or clinic. An inclination or natural tendency to behave in a particular way. Within EDH, the over 40 propensity scores calculated relate to different inpatient and outpatient services and is a determination of how likely a person is to need services related to a medical service line. Examples include: ip – obstetrics, ip – orthopedics, op – plastic surgery, op – urology, etc.
Recipe	<p>Default criteria used to start building a list of persons for a specific type of marketing campaign.</p> <p><b>Example:</b> a mammography recipe can be used to quickly start a list for a marketing campaign to remind women to get a mammography when the time is right.</p> <p>A calculated score based on a set of filters that indicate a person's targeting value for a campaign. Recipes can be based on demographic, utilization, financial, and propensity information. Current examples include: Allergy Symptoms, Weight Management Services, Orthopedics – General, etc.</p>
Redshift	The EDH data warehouse, hosted by AWS. ETL jobs are run against Cassandra to move the data to Redshift for access from Aginity to run queries and produce off line list pulls.
S3	Amazon Simple Storage Service, Amazon S3 is <a href="#">object storage</a> built to store and retrieve any amount of data from anywhere – web sites and mobile apps, corporate applications, and data from IoT sensors or devices. It is designed to deliver 99.99999999% durability.
Service Area	And Internal term used to describe a list of Zipcodes that comprise the area for which a hospital/clinic serves. A hospital or clinic might market to persons within this area. Most list are created from a specific service area. Service areas are generally grouped into a primary service area, a secondary service area, and a tertiary service area.
Spark	Apache Spark, open-source cluster-computing framework. Most of the EDH is scripted Spark jobs for the manipulation of CRM data.